# Exploratory Data Analysis of Suicide Rates

(A COMPREHENSIVE DATA ANALYSIS & STATISTICAL APPROACH)

# Introduction to Suicide Analysis

- Suicide is a major public health concern worldwide, influenced by economic, social, and psychological factors.
- This study aims to analyze suicide trends, key contributing factors, and their statistical significance.
- Techniques used: Data preprocessing, Exploratory Data Analysis (EDA), Statistical Analysis, and Hypothesis Testing.

# Understanding the Dataset

- **Columns included in the dataset:**

1. Country
2. Year
3. Gender
4. Age Group
5. Suicide Count
6. Population
7. GDP per Capita
8. Human Development Index (HDI)

Objective: Identify trends and key predictors of suicide rates.

# Project Workflow

- **Data Cleaning:** Handling missing values, identifying and managing outliers, normalizing data.
- **Exploratory Data Analysis (EDA):** Visualization using Matplotlib & Seaborn, correlation & covariance analysis.
- **Statistical Analysis:** Descriptive & inferential statistics, hypothesis testing.
- **Key Findings:** Identifying the top factors contributing to suicide rates.

# Data Cleaning

- **Handling Missing Values:**
  - Missing values were identified using .isnull() .sum().
  - Imputed missing values using mean for numerical columns.
- **Identifying Outliers:**
  - Box plots were used to detect extreme values.
  - Outliers were not removed because removing them could distort meaningful patterns in the data.
- **Normalization:**
  - Used MinMaxScaler to scale numerical features for better comparability

# Exploratory Data Analysis (EDA)

- **Histogram Analysis:**
  - Most features showed a right-skewed distribution.
  - Suicide rates varied significantly based on sex, generation, and age group.
- **Boxplot Analysis:**
  - Gender-based differences in suicide rates were clearly evident.
- **Scatter Plot Analysis:**
  - GDP per capita showed a moderate correlation with suicide rates.
- **Line Plot Analysis:**
  - Suicide rates fluctuated over time, with noticeable peaks in certain years

# Correlation & Covariance Analysis

- **Correlation helps measure the strength of relationships between variables:**
  - GDP per capita had a moderate correlation with suicide rates.
  - Population size was strongly correlated with absolute suicide numbers.
- **Covariance measures the direction of relationships between variables:**
  - Economic factors like GDP & HDI showed positive covariance with suicide rates.

# Statistical Analysis

- **Measures of Central Tendency:**
  - Mean, median, and mode of suicide rates were analyzed across different demographics.
- **Measures of Dispersion:**
  - Variance and standard deviation were calculated to assess variability.
- **Inferential Statistics:**
  - T-Test and ANOVA were conducted to identify significant relationships.

# Hypothesis Testing

- **T-Test:** Checked if the mean suicide rate of a sample significantly differed from the population mean.
- **T-Test: Checked if the mean suicide rate of male and female defers alot**
- **ANOVA:** Analyzed the effect of different age groups on suicide rates.

- **Findings:**

1. No significant difference was found between sample and population suicide rates.
2. GDP per capita was significantly associated with suicide rates.
3. Suicide rates varied significantly across different age groups.

# Top Factors Contributing to Suicide Rates

1. **GDP per Capita:** Economic instability increases suicide risk.
2. **Age & Gender:** Older individuals and males show higher suicide rates.
3. **Health & Life Expectancy:** Poor mental health and limited healthcare access contribute to higher suicide risks.
4. **Human Development Index (HDI):** Countries with low HDI show higher suicide rates due to poor healthcare, education, and income levels.

# Key Insights

- **Economic Stability Matters:** Countries with higher GDP per capita generally have lower suicide rates.
- **Demographics Play a Role:** Older individuals and males are at higher risk.
- **Healthcare Access is Crucial:** Life expectancy and mental health services significantly impact suicide rates.
- **HDI is a Strong Indicator:** Low HDI nations require better social and economic policies to reduce suicide rates.

# Conclusion

- **Suicide rates** are influenced by a combination of **economic, social, health, and demographic factors**.
- Economic instability, lack of healthcare access, and societal pressures contribute to higher suicide risks.
- Governments and policymakers should focus on **mental health awareness, economic stability, and healthcare improvements** to reduce suicide rates.